In the Claims:

1. A method for estimating a selectivity of a query containing at least one column-associated condition related to column attributes of a relational database table, the method comprising:

(a) generating a dataset by sampling a plurality of queries applied against the database, wherein the dataset includes a plurality of query conditions and information related to combinations of said query conditions,

(b) determining at least one regression function that reflects correlations between particular query conditions based on said dataset,

(c) determining a table-specific estimate of a cardinality of a query based upon the regression function serving as a data mining model.

2. The method of claim 1, wherein step (c) further includes:

(c.1) selecting an access method for an incoming query from a plurality of database access methods based upon the table-specific estimate for said incoming query.

3. The method of claim 1, wherein said query includes column associated conditions related to a plurality of tables, wherein step (c) further includes:

determining a table-combining cardinality estimate based upon said table-specific estimate.

4. The method of claim 1, wherein step (a) further includes:

(a.1) generating a dataset including queries q_j, j= 1,..N, wherein each query includes a plurality of column-associated conditions c_{jk}, k= 1,..M_j, N, M being integer variables, wherein step (a.1) further includes:

(a.1.1) storing a cardinality C of an elementary operation associated with a column-associated condition c_{jk},

(a.1.2) storing a count of query-qualifying database records reflecting the correlation between the database table column attributes referred to in each

elementary operation,

wherein step (c) further includes:

(c.1) calculating a cardinality estimate CE of said query with the following formula:

$$CE = \sum_{i=1,..L} f(Z_i)$$

wherein $f(Z_i)$ is a regression function, CE is a total of correlations between the plurality of combinations of elementary operations used in said sampled queries, and Z_i is a frequency of occurrence for one or more column-associated conditions c_{jk}, and

wherein step (b) further includes:

(b.1) generating said regression function using said data mining model.

5. The method of claim 4, wherein step (c) further includes:

(c.2) estimating the cardinality of each of the plurality of column-associated conditions c_{jk} referring to the same column using the data mining model.

6. The method of claim 1, wherein step (c) further includes:

(c.1) training the model by using queries that include logical AND operators to determine a correlation between corresponding column predicates.

7. The method of claim 1, wherein step (c) further includes:

(c.1) transforming a query containing OR predicates to an equivalent query containing AND predicates to simplify training of a model.

8. The method of claim 1, wherein step (c) further includes:

(c.1) normalizing the determined cardinality based upon a total number of rows in the database table.

9. The method of claim 1, wherein step (c) further includes:

(c.1) normalizing the cardinality associated with a sampled query with a size of the database table when the query is sampled, and

(c.2) denormalizing a cardinality associated with a query for which a

cardinality is to be predicted with the size of the database table when the selectivity for that query is predicted.

10. The method of claim 1, wherein step (b) further includes:

(b.1) using a subset of frequently used queries to determine said regression function.

11. The method of claim 1, wherein step (b) further includes:

(b.1) repeatedly training said regression function with updated sampled data.

12. The method of claim 1, wherein step (a) further includes:

(a.1) sampling said queries via a tool based on a database optimizer.

13. The method of claim 1, wherein step (a) further includes:

(a.1) determining cardinalities for individual table columns via a database statistics tool, and

(a.2) mapping queries that include a plurality of logical AND operators to corresponding cardinality based regression formulae.

14. The method of claim 1, wherein step (a) further includes:

(a.1) mapping queries that include at least one of an inner join and an outer join to corresponding regression formulae based on at least one of cardinality and selectively operations.

15. A method for determining an access plan for a database query compatible with data mining based database access control comprising:

(a) selecting a regression function for use with said query,

(b) determining a number of qualifying records for said query via said regression function, and

(c) selecting an access method for accessing the database from a plurality of different access methods based upon the determined number of qualifying records.

16. A computer system for estimating a selectivity of a query containing at least one column-associated condition related to column attributes of a relational database table, the system comprising:

a sampling module for generating a dataset by sampling queries applied against the database, wherein the dataset includes a plurality of query conditions and information related to combinations of said query conditions,

a regression module for determining at least one regression function that reflects correlations between particular query conditions based on said dataset,

a processing module for determining a table-specific estimate of a cardinality of a query based upon the regression function serving as a data mining model.

17. The system of claim 16, wherein the processing module selects an access method for an incoming query from a plurality of database access methods based upon the table-specific estimate for said incoming query.

18. The system of claim 16, wherein said query includes column associated conditions related to a plurality of tables, and wherein the processing module determines a table-combining cardinality estimate based upon said table-specific estimate.

19. The system of claim 16, wherein the sampling module further comprises:

a dataset module for generating a dataset including queries q_j, j= 1,..N, wherein each query includes a plurality of column-associated conditions c_{jk}, k= 1,..M_j, N, M being integer variables, wherein said dataset module further comprises:

a first storage module for storing a cardinality C of an elementary operation associated with a column-associated condition c_{jk},

a second storage module for storing a count of query-qualifying database records reflecting the correlation between the database table column attributes

referred to in each elementary operation,

wherein the processing module further comprises:

an estimation module for determining a cardinality estimate CE of said query with the following formula:

$$CE = \sum_{i=1,..L} f(Z_i)$$

wherein $f(Z_i)$ is a regression function, CE is a total of correlations between the plurality of combinations of elementary operations used in said sampled queries, and Z_i is a frequency of occurrence for one or more column-associated conditions c_{jk}, and

wherein the regression module further comprises:

a function module for generating said regression function using said data mining model.

20. The system of claim 19, wherein the processing module estimates the cardinality of each of the plurality of column-associated conditions c_{jk} referring to the same column using the data mining model.

21. The system of claim 16, wherein the processing module trains the model by using queries that include logical AND operators to determine a correlation between corresponding column predicates.

22. The system of claim 16, wherein the processing module transforms a query containing OR predicates to an equivalent query containing AND predicates to simplify training of a model.

23. The system of claim 16, wherein the processing module normalizes the determined cardinality based upon a current total number of rows in the database table.

24. The system of claim 16, wherein the processing module normalizes the cardinality associated with a sampled query with a size of the database table when the query is sampled, and denormalizes a cardinality associated with a query for which a cardinality is to be predicted with the size of the database table when the selectivity for that query is predicted.

25. A program product apparatus having a computer readable medium with computer program logic recorded thereon for estimating a selectivity of a query containing at least one column-associated condition related to column attributes of a relational database table, said program product apparatus comprising:

a sampling module for generating a dataset by sampling queries applied against the database, wherein the dataset includes a plurality of query conditions and information related to combinations of said query conditions,

a regression module for determining at least one regression function that reflects correlations between particular query conditions based on said dataset,

a processing module for determining a table-specific estimate of a cardinality of a query based upon the regression function serving as a data mining model.

26. The program product of claim 25, wherein the processing module selects an access method for an incoming query from a plurality of database access methods based upon the table-specific estimate for said incoming query.

27. The program product of claim 25, wherein said query includes column associated conditions related to a plurality of tables, and wherein the processing module determines a table-combining cardinality estimate based upon said table-specific estimate.

28. The program product of claim 25, wherein the sampling module further comprises:

a dataset module for generating a dataset including queries q_j, j= 1,..N, wherein each query includes a plurality of column-associated conditions c_{jk}, k= 1,..M_j, N, M being integer variables, wherein said dataset module further comprises:

a first storage module for storing a cardinality C of an elementary operation associated with a column-associated condition c_{jk},

a second storage module for storing a count of query-qualifying database records reflecting the correlation between the database table column attributes referred to in each elementary operation,

wherein the processing module further comprises:

an estimation module for determining a cardinality estimate CE of said query with the following formula:

$$CE = \sum_{i=1,..L} f(Z_i)$$

wherein $f(Z_i)$ is a regression function, CE is a total of correlations between the plurality of combinations of elementary operations used in said sampled queries, and Z_i is a frequency of occurrence for one or more column-associated conditions c_{jk}, and

wherein the regression module further comprises:

a function module for generating said regression function using said data mining model.

29. The program product of claim 28, wherein the processing module estimates the cardinality of each of the plurality of column-associated conditions c_{jk} referring to the same column using the data mining model.

30. The program product of claim 25, wherein the processing module trains the model by using queries that include logical AND operators to determine a correlation between corresponding column predicates.